School of Built Environment, Engineering & Computing, Leeds Beckett University, Leeds LS1 3HE, UK
Department of Computer Science, Kinnaird College for Women, Lahore 54000, Pakistan
Author to whom correspondence should be addressed.

Sensors **2023**, *23*(21), 8972; https://doi.org/10.3390/s23218972

Received: 26 September 2023 / Revised: 27 October 2023 / Accepted: 2 November 2023 / Published: 4 November 2023

(This article belongs to the Section Internet of Things)

Download Browse Figures Versions Notes

Abstract

The Internet of bio-nano things (IoBNT) is an emerging paradigm employing nanoscale (~1–100 nm) biological transceivers to collect in vivo signaling information from the human body and communicate it to healthcare providers over the Internet. Bio-nano-things (BNT) offer external actuation of in-body molecular communication (MC) for targeted drug delivery to otherwise inaccessible parts of the human tissue. BNTs are inter-connected using chemical diffusion channels, forming an in vivo bio-nano network, connected to an external ex vivo environment such as the Internet using bio-cyber interfaces. Bio-luminescent bio-cyber interfacing (BBI) has proven to be promising in realizing IoBNT systems due to their non-obtrusive and low-cost implementation. BBI security, however, is a key concern during practical implementation since Internet connectivity exposes the interfaces to external threat vectors, and accurate classification of anomalous BBI traffic patterns is required to offer mitigation. However, parameter complexity and underlying intricate correlations among BBI traffic characteristics limit the use of existing machine-learning (ML) based anomaly detection methods typically requiring hand-crafted feature designing. To this end, the present work investigates the employment of deep learning (DL) algorithms allowing dynamic and scalable feature engineering to discriminate between normal and anomalous BBI traffic. During extensive validation using singular and multi-dimensional models on the generated dataset, our hybrid convolutional and recurrent ensemble (CNN + LSTM) reported an accuracy of approximately ~93.51% over other deep and shallow structures. Furthermore, employing a hybrid DL network allowed automated extraction of normal as well as temporal features in BBI data, eliminating manual selection and crafting of input features for accurate prediction. Finally, we recommend deployment primitives of the extracted optimal classifier in conventional intrusion detection systems as well as evolving non-Von Neumann architectures for real-time anomaly detection.

Keywords: deep learning; internet of bio-nano things; bioluminescence; bio–cyber interface; intrusion detection; anomaly detection; traffic classification; neuromorphic computing

1. Introduction

The Internet of Things (IoT) presents a connected environment where real-life objects like sensors, actuators, and electronic devices can interact with each other through enabling technologies such as the Internet. The concept of IoT has recently been revised in the light of novel nanotechnology tools and synthetic biology, which have resulted in the fabrication of biologically embedded computing devices at the scale of nanometers (1–100 nm) called Bio-Nano things (BNT) [1]. The minute size puts constraints on their operational resources like computational complexity and storage capabilities. Therefore, bio-nano things are only able to perform trivial tasks like sensing, data storage, and actuation. The capabilities of a single BNT device, in terms of complexity and range of operation, can be expanded when allowed to interact with counterparts forming a network called bio-nanonetwork (BNN). Nanonetwork enables bio-nano things to share, fuse, and coordinate their information in the biological environment through a novel communication technology called Molecular Communication (MC) [2]. MC enables the biochemical processing of data, the transformation of chemical energy, and the exchange of information through the transmission and reception of molecules. These characteristics of MC can be advantageous for environments like the human body, where the use of electromagnetic (EM) radiation for healthcare could result in unwanted effects on human tissue. The connection of MC to external networks such as the Internet leads to a novel communication paradigm called the Internet of Bio-Nano Things (IoBNT) [1,3]. IoBNT opens up a plethora of biomedical applications such as real-time intra-body sensing and actuation, targeted drug delivery (TDD), and tissue re-engineering [4]. IoBNT applications, allow TDD and remote patient monitoring, in the ease of patients' home or workplace by the healthcare provider. It can, therefore, be inferred that IoBNT interfaces will aid in connecting the biological environment of the human body to the electrical cyber world. In this scenario, where the human body can be monitored and injected with drugs remotely, the security of the users is of paramount importance. Several cyber-attacks and viruses can be launched on the bio-cyber interface to disrupt in vivo MC-related information being sent to the healthcare provider, as well as alter or disrupt physiological functions and homeostasis of the body, giving birth to a new form of terrorism called *bio-cyber terrorism* [1,4]. Among the existing proposals for bio-cyber interfacing technologies, bioluminescent-based interfaces (BBI) have demonstrated realistic incorporation in the IoBNT paradigm due to their low cost and highly efficient operation [4,5]. Securing BBI against inherent threat vectors requires state-of-the-art intrusion detection systems (IDS). This work proposes a comprehensive anomaly detection framework for bioluminescence-based bio–cyber interface (BBI) in IoBNT systems using deep learning classification. A combination of Deep Learning-Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) based profiling scheme is used to differentiate the anomalous activity of bio–cyber interface. In the case of traffic classification, CNN alone cannot perform well as it does not specifically account for temporal characteristics. In a malicious attack, several traffic modalities can be altered in a short period of time to modify in-vivo bio-interface parameters. In this situation applying CNN alone may lead to missed alerts. Therefore, CNN is combined with LSTMs, which support sequence prediction and account for time-related characteristics of abnormal operations. The hybrid CNN-LSTM architecture involves using CNN layers for feature extraction on input data and LSTM to analyze sequence prediction and temporal information. The proposed security framework is trained using extensive simulations on data presenting significant attributes of bioluminescence-based bio–cyber interface. We address the system engineering perspective by proposing and comparing different scenarios for implementing deep learning (DL) classifiers for anomaly detection in IoBNT systems. The rest of this paper is organized as follows. Section 2 highlights the background on the Internet of Bio Nano things and bio–cyber interfacing, presents related work, and presents applications of deep learning in cyber security. Section 3 discusses bio–cyber interfacing, and measurement parameters to be used for anomaly detection. Section 4 discusses the proposed methodology, data collection and pre-processing, and the neural network structures used. Section 5 evaluates the results from a qualitative and quantitative perspective and presents system deployment scenarios for the derived classifier. Final conclusions are drawn in Section 6.

2. Background

In the present section Internet of Bio Nano Things, related work, and an overview of deep learning techniques in network security are summarized.

2.1. Internet of Bio Nano Things-Overview

The architecture of IoBNT is a holistic framework that encapsulates heterogeneous devices (nano-macro scale) and communication protocols (nano communication to macro-scale communication). A reference architecture of IoBNT is presented in Figure 1. The IoBNT paradigm was first proposed by Akyildiz et al. [1] and followed by further research contributions in the design and technological aspects of IoBNT such as bio–cyber interfacing by Chude-Okonkwo et al. [5] and Nakano et al. [2], and molecular communication primitives by Nakano et al. [6,7] and and Felicetti et al. [8]; nanonetworks by Akyildiz et al. [3]; communication channel characteristics by Garralda et al. [9], Kuran et al. [10], and Gregori and Akyildiz [11]. The basic unit of IoBNT is the Bio-Nano Thing (BNT). Bio-Nano things are often referred to as nano things, nanomachines, and nano devices in literature. The size of bio nano things is of the order of nanometres (100 nm). Bio-Nano things can be composed of only biological materials (lipids, proteins, etc.) or can be coated with non-biological materials like magnetic particles. The design of bio-nano things can be customized according to the task at hand, e.g., liposomes are capable of storing and releasing molecules, long circulation, targeting, and stimuli responsiveness [12,13]. Genetically engineered biological cells, bacterium, and artificial cells have the capabilities to synthesize and emit specific molecules and thus can also be used as bio-nano things [5]. To equip the bio-things with sensing abilities, a special nanosensor can be designed using a whole cell [14,15] and sensor molecule entrapped in a chemically inert matrix [16]. All the above-mentioned bio-nano things collaboratively work in the nano-networking setup. Bio-nano things traverse the blood vessel network and transport the encoded molecular information to the bio–cyber interface. The design of Bio–cyber interface is an open research issue in the realization of the IoBNT paradigm. Research efforts have been devoted towards interfacing technologies such as electronic tattoos, hydro-gel conductors, and chemo-electro transduction units to offer connectivity between the human body and the Internet with considerable success [5,17,18,19]. Among the existing interfacing technologies bioluminescent interfacing has provided promising results in practical implementation owing economy of cost and ease of deployment [4,5,7,9].

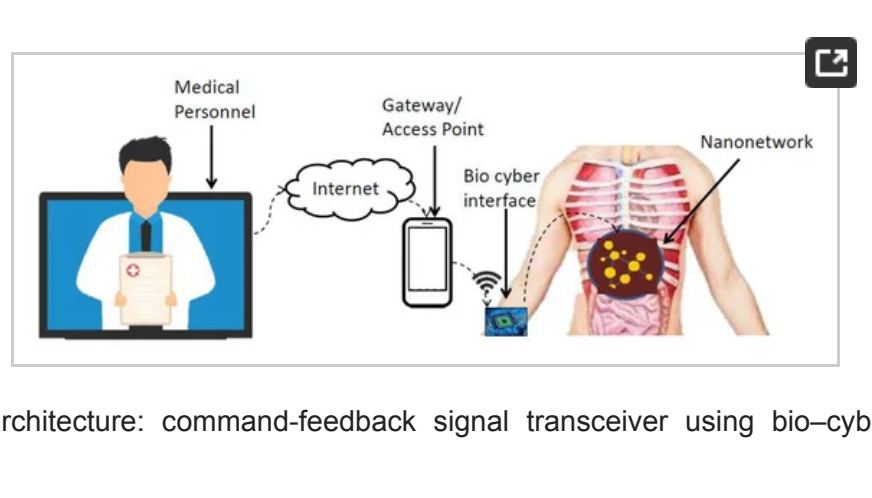

Figure 1. IoBNT architecture: command-feedback signal transceiver using bio–cyber interface over the Internet.

2.2. Related Work

The realization of a secure bio–cyber interface will aid in wider adoption of the promising healthcare applications of IoBNT. Bioluminescence-based IoBNT is moving towards greater realization, requiring security primitives able to monitor and identify in vivo traffic anomalies triggered by external (Internet) cyber-attacks [4]. A brief overview of the state-of-the-art IoBNT security and complementary research in associated domains is presented in Table 1. Pioneering surveys such as [1,2,4,6] have primarily focused on systematically reviewing bio–cyber interfacing technologies and documenting underlying security requirements. Research contributions in formulating a holistic intrusion detection and prevention (IDS, IPS) framework for IoBNT are, however, still nascent. In the direction of molecular communication (MC), some preliminary studies have been conducted to explore the security requirements of nanonetworks, attack categorization, as well as authentication mechanisms [20,21]. Giaretta et al. [22] classified different attack types (black hole and sentry) along with their countermeasures for BNT security. Zafar et al. [23] proposed an authentication scheme for diffusion-based MC using Channel Impulse Response (CIR) as a device fingerprint. Recent studies have sought to use machine learning-based approaches to classify and discriminate network traffic for bio–cyber interface security. El-Fatyany et al. [24] presented an authentication scheme for bio luminescence-based bio–cyber interface (BBI). The proposed chaotic model utilizes a modified logistic map to control drug release in a targeted drug delivery system. Similarly, an earlier work, by Bakhshi et al. [25] devised an authentication scheme for bio–cyber interface technologies using ML-enabled profiling. The security framework was designed for three types of bio–cyber interfacing technologies namely bioFET, bioluminescence, and reduction-oxidation (redox) modality. However, the study focused on using only a limited number of external traffic parameters as input features for ML training. While ML-based profiling utilizing manually crafted input parameters have proven to be somewhat efficient in traditional IDS (and IPS), the inter-dependence and inherent complexity of bio-nano traffic constrains their optimality for anomaly detection in the IoBNT domain. Increasing levels of data (traffic) intricacy and a requirement for greater classification accuracy have led researchers to investigate applications of deep learning neural networks in the mainstream IoT domain. DL network structures have been shown to aid dynamic feature selection and significantly reduce false positives during traffic classification in heterogeneous IoT settings [26,27]. Abusitta et al. [26] highlighted a preference for DL-based classification to recognize abnormal behavior in IoT traffic, prone to external noise over common ML algorithms. Using de-noising autoencoders during the data pre-processing stage, the DL classifier was able to discriminate malicious traffic patterns by dynamically extracting useful features despite an unstable operational environment. Similarly, Rosero-Montalvo et al. [27] used a hybrid DL ensemble to detect anomalies with minimal memory and computational footprint in high-noise IoT environments. Real estate and workload economy enabled DL classifier realization on off-shelf IoT sensory hardware as well as ensured high bandwidth efficiency. Chander et al. [28] focused on industrial IoT (IIoT) sensors, developing metaheuristic feature selection, and tested single and multidimensional cascaded recurrent neural networks for the identification of anomalies. The resulting classifier allowed satisfactory anomaly detection accuracy which was further improved by model tuning. Hameed et al. [29] compared the efficacy of existing ML solutions for IoT traffic classification in smart cities. The authors were of the view that fine-grained flow characteristics and correlations among IoT traffic parameters presented a computational concern in typical ML-based classifiers. Using multi-layer perceptron—artificial neural networks (MLP-ANN), and a two-stage learning framework researchers were able to achieve high accuracy and precision in identifying IoT devices based on network traffic, significantly outperforming well-known ML algorithms. Guan J. et al. [30] highlight the scarcity of IoT labeled data in 5G environments, the limited capability of devices, and the significant efforts involved in artificially designing features for traffic classification. Employing deep transfer learning, the researchers used customized convolutional networks trained on resource-constrained targets to provide strong traffic classification accuracy. Reducing human intervention during feature selection, the framework decreased system complexity for future adoption in 5G IoT networks to isolate traffic of interest. Etemadi et al. [31] surveyed abnormality detection at the molecular communication level and compared different ML, heuristics, and DL primitives for this purpose. The work discussed the viability of ML approaches, as well as the utilization of naturally inspired systems based on genetic, particle swarm, and colony optimization algorithms for identifying anomalous (traffic) activity. It was observed that while these techniques are appropriate for stationary sensors and objects, it became very challenging to address moving sensors and abnormality sources. The authors recommended using a combination of heuristic deep learning methods and localized customization to identify anomalous behavior in IoBNT. The application of DL techniques for anomaly detection in network engineering, however, remains prevalent [32,33]. Preliminary efforts in using DL-based anomaly detectors for bioFETs have been investigated by Zafar et al. [33]. The authors proposed an Artificial Neural Network-based classifier for biological field effect transistor (bioFET) based bio–cyber interface. Particle Swarm Optimization (PSO) was used to adjust the weights and biases of the neural networks to provide better classification. The proposed model utilizes communication and design parameters of bioFET-based bio–cyber interface to classify normal versus abnormal traffic behavior. The present work seeks to expand on the utility of DL-based network structures through parameter profiling of bioluminescent interfaces and allowing timely detection of anomalies to realize secure avenues for IoBNT deployment in future healthcare applications.

Table 1. Related work and cross-connecting themes.

2.3. Deep-Learning Enabled Traffic Analysis

Traditional machine learning techniques including Iterative Dichotomiser 3 (ID3), decision tree (C 5.0), classification and regression tree (CART), and artificial neural network (ANN) have been previously recommended to provide a measure of security for IoBNT biocyber interfaces [25,26]. Deep learning is a sub-field of traditional machine learning, providing increased flexibility and accuracy over classical learning algorithms, and offering the ability to incrementally learn and extrapolate new features from a limited set of training data [34,35]. Moreover, the thin and layered structure of sequential deep neural network models makes them ideal for being deployed over a low-powered and resource-constrained bio–cyber interface, still facilitating real-time anomaly detection. Two of the popular deep learning network types being used in several image recognition, video editing, industrial, and security systems include the convolutional and recurrent neural networks [34].

Convolutional neural network (CNN) is an extension of the traditional feed-forward neural network. CNN has been proven significantlyly successful in image processing tasks including object identification in robotics, and self-driven cars, and are now gaining traction in cyber security application particularly anomaly detection [32,34,35,36]. Four main operations of CNN comprise of a convolution layer responsible for feature extraction, a non-linear activation function such as Rectified Linear Unit (ReLu) for fast(er) model learning, a pooling layer for generalization, and a fully connected layer (to be used for classification) [37]. The number of neurons in hidden layers relates to model learning ability, albeit at the cost of computation. CNNs are extremely useful in image processing applications, however, usually offer limited capability in effectively interpreting temporal data characteristics inherent in cyber security primitives, traffic classification, and intrusion detection systems [36].

Recurrent neural networks (RNN) structures are used for processing time-series data by employing neurons capable of remembering previously processed (data) states [38]. RNNs are based on human-cognitive functionality and can generate input-output mappings considering time-dimensionality (i.e., order of occurrence). The network structure is based on feedback loops, using the previous state to compute the next state. For longer data sequences, RNNs however, suffer from a vanishing gradient problem (short-term memory) resulting in high latency and signal transfer deviation from one (prior) iteration to the next. To deal with vanishing gradient and to address memory problems, variations in RNN including the long-short term memory (LSTM) and gated recurrent unit (GRU) have been proposed [39].

Long short-term memory (LSTM) is a refined extension of RNNs, developed to reduce the gradient explosion and dispersion issue. LSTM introduces memory blocks, in contrast to the conventional simple RNN units. A memory block contains a memory cell and a set of gates i.e., input, forget, and output. This has emerged to be more effective in sequence information analysis and capturing long-range dependencies with applications in network security [35]. LSTM filters prior states based on greater influence on the present.

Gated recurrent unit (GRU) is also a type of RNN comprising of an update and reset gate, trained to select, and retain information that needs to be fed to the output. The update gate allows the determination of past information to be sent to future states fully or partially and significantly reduces the vanishing gradient problem. The reset gate determines the scope of forgetting past information. While the applicability of CNN and RNN structures is widespread in image processing and handwriting recognition, their applicability in anomaly detection systems is relatively nascent [4,24]. To the best of our knowledge, the present work is the first to consider deep learning structures for the security of IoBNT devices and interfaces.

3. Bioluminescence Based Bio–Cyber Interface Architecture

Bio–cyber interface is a hybrid device capable of transducing in-body biochemical signals into electromagnetic signals and vice-versa. Several modalities can be used for transduction between biochemical and electromagnetic signals such as bioFET-based molecular antenna [17], reduction-oxidation (Redox) modality [18,19], and other interfaces offering biologically inspired capabilities like thermo-responsiveness and photo-responsiveness [5]. The choice of interfacing modality depends upon the assigned application requirements and propagation channel [4]. In this work, we have considered the bioluminescence-based biocyber interface model proposed by Chude-Okonkwo et al. [5] based on biologically inspired properties such as biocompatibility, biodegradability, and non-toxicity. The bioluminescent biocyber interface (BBI) provides a practical opportunity for healthcare professionals to connect the Internet to in-body (in vivo) bio nanonetworks and studies in the domain are maturing to prototype evaluations [40]. We present a sample architecture and bio-cyber interface model illustrated in Figure 2. The bio-cyber interface utilizes thermal and light responsiveness to certain biomolecules producing information signals (biochemical reactions) propagated using the blood vessel network to an in vivo nanonetwork location for targeted drug delivery. The architecture of the bio–cyber interface includes two units: an electro-bio transduction unit and a bio-electro transduction unit. A detailed description of these units is given below.

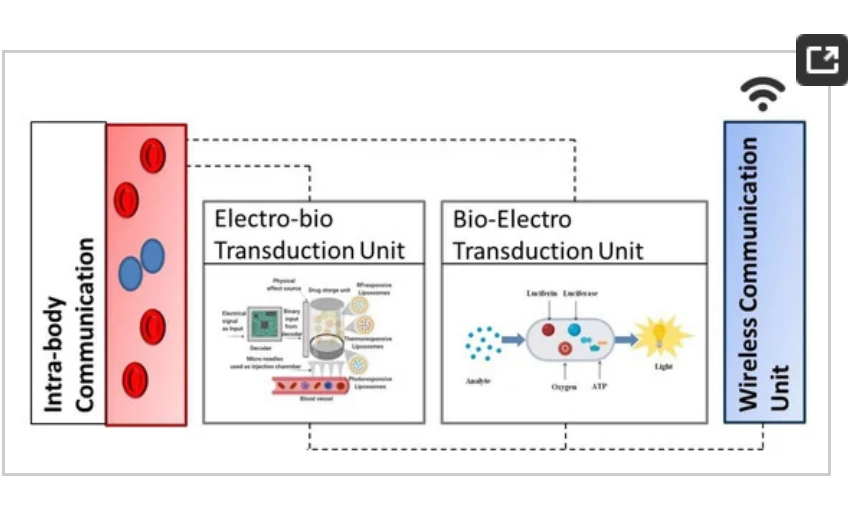
Intra-body Communication
Electro-bio Transduction Unit
Bio-Electro Transduction Unit
Wireless Communication Unit

Figure 2. An illustration of a generic bio–cyber interface.

3.1. Electro-Bio Transduction Unit

Electro-bio transduction unit (EBTU) converts electrical signals/commands received from medical personnel into biochemical signals. A visual illustration of EBTU is provided in Figure 3.

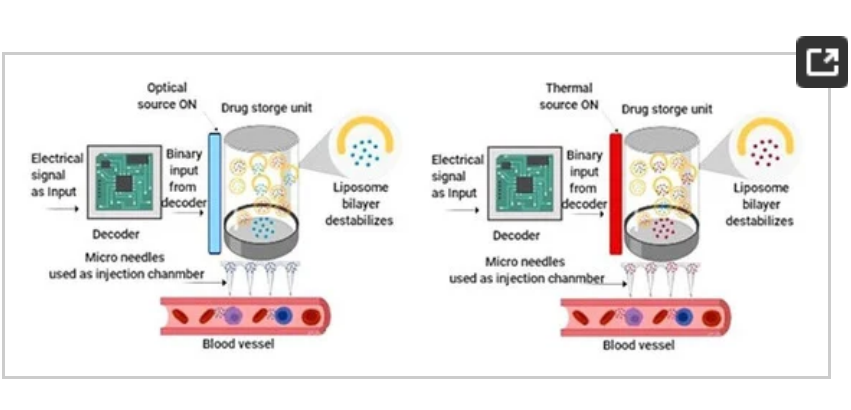
Optical source ON
Drug storge unit
Electrical signal as input
Binary input from decoder
Liposome bilayer destabilizes
Decoder
Micro needles used as injection chanmber
Blood vessel
Thermal source ON
Drug storge unit
Electrical signal as input
Binary input from decoder
Liposome bilayer destabilizes
Decoder
Micro needles used as injection chanmber
Blood vessel

Figure 3. Electro-bio transduction unit on left when the optical source is ON and on the right when the thermal source is ON.

The command received from external devices is in the form of a binary code. This binary code is used to drive logic gates to produce an optical or thermal effect. The unit also contains a small chamber/drug reservoir that contains nano-sized (1–100 nm) carriers. These nanocarriers are designed in such a way that they release their contents upon stimulation from external factors such as changes in temperature, light, pH, etc. The bio–cyber interface model based on [5] adopted for this work employs two types of liposomes: *photo-responsive* liposomes and *thermo-responsive* liposomes. The photo-responsive release is considered through *photoisomerization*, where liposomes encapsulate molecules that excite upon the light illumination from an external source, this causes a conformational change and destabilization of the lipid membrane that allows molecule release. For thermo-responsiveness, molecules are encapsulated in temperature-sensitive liposomes or dendrimers [41], which deteriorate upon receiving nonlinear sharp changes in the temperature. To preserve the contents during propagation, thermo-responsive liposomes must retain their load at body temperature (37 °C) and may release their contents within a locally heated microenvironment. The release process of liposomes can be expressed in the following equation:

$$v(t) = \varepsilon_T\,(1 - e^{(\gamma t)}) \tag{1}$$

where ε_T is the cumulative molecular concentration and γ is the release rate of liposomes. The output of the electro-bio unit can be expressed as follows.

$$c^{(f)} = \int_{(0\rightarrow T_{in})} \varepsilon\, \Psi\,(t)\, dt \tag{2}$$

where T_{in} is the time difference between injection and the start of the release process and Ψ is the total number of liposomes that release their content. The communication path of the electro-bio transduction unit is from bio–cyber interface towards in-body nanonetworks. The concentration of molecules may change during the propagation time i.e., after being injected from the bio–cyber interface into the blood vessel network and reaching the nano network or designated tissue. This rate of change of molecular communication from electro-bio transduction unit towards nanonetwork can be modeled through the equation below [5]:

$$d_t v_1(t) = v_1(t)\,(k_{12} + k_{10}) + k_{21}\, v_2(t) \tag{3}$$

$$d_t v_2(t) = k_{12}\, v_1(t) - k_{21}\, v_2(t) \tag{4}$$

with initial conditions $v_1(0) = c^{(f)}$ and $v_2(0) = 0$. Where k_{12} and k_{21} are first-order rate constants, k_{10} is the elimination rate, $v_1(t)$ is the molecular concentration in the blood vessel network, and $v_2(t)$ is the concentration of information molecules when they reach nanonetwork. k_{10} is the elimination rate and $v_{el}(t)$ is the function of k_{10}. The elimination rate represents the number of molecules that undergo biochemical modification, phagocytosis, elimination by liver or adhesion, and absorption by non-targeted sites, during the propagation process. Generally, the rate constants depend on the concentration difference between the blood vessel network and nanonetwork, the size of the aperture through the endothelial cell network, and the properties of the diffusing information molecules [42].

3.2. Bio-Electro Transduction Unit

This unit presents reverse communication, i.e., detection of biochemical signals in the blood vessel network and converting it to an equivalent electromagnetic signal. A visual illustration of bioreporter activity is presented in Figure 4.

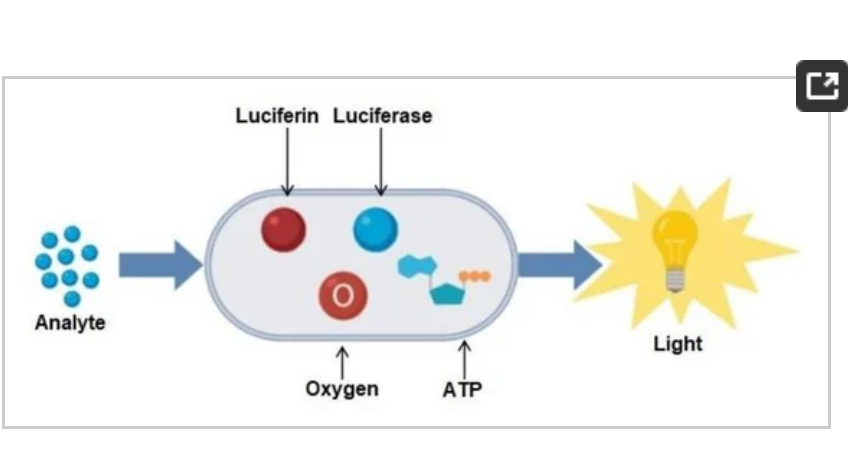

Figure 4. Representation of the bioreporter as a cellular structure, in which information molecules diffuse and a reporter protein (LU) is produced and undergoes a bioluminescence reaction in the bioreporter.

This unit works on the principle of bioluminescence reaction. The conceptual model of this unit is a ligand–receptor transdermal system whose nanopores are extended into the blood vessels. The ligands here can be thought of as information molecules that circulate the blood vessel network and are detected by complementary receptors/nanopores. This ligand–receptor pairing results in the production of an equivalent electrical signal. Typically, luminescent materials are used to convert chemical signals into optical signals so that they can be detected by external devices. Bioluminescence is a phenomenon where a chemical reaction produces an electronically excited substance, called an enzyme-substrate system. The enzyme is Luciferase (LU), and the substrate is Luciferin(L) (from Lucifer in Latin "Light-bringer") [43]. Bioluminescent molecules such as LU catalyze a chemical reaction with L in the presence of oxygen, which consumes chemical energy e.g., Adenosine triphosphate (ATP). ATP is a central metabolite that plays the role of an energy transfer molecule, a phosphate donor, as well as a signaling molecule inside the cells and produces light. The resultant light in turn is detected by a highly sensitive light sensor and is transduced into an electrical signal afterward. Bioluminescence intensity factor *I* can be expressed in the form of the Michaelis-Menten mechanism as follows [5]:

$$I(t) = a_l.l_a.atp/(atp + a_M) \tag{5}$$

where a_l is the catalytic reaction, l_a is the concentration of LU, atp is the concentration of ATP and a_M is the Michaelis-Menten constant. Let I_0 be the bioluminescence intensity, hence when $I(t) \geq I_0$ the transmitter is in the ON state, and, if $I(t) \leq I_0$ the transmitter is in the OFF state. Bio-electro transduction unit receives signal from nanonetworks in the form of molecular concentration. The concentration of information molecules may change during the propagation time i.e., from nanonetwork/designated tissue into the blood vessel network and reaching the bio–cyber interface. This rate of change of molecular communication of the bio-electro transduction unit can be expressed through the equations below [5]:

$$c_{(r)}\,(t) = k_l\, w_1\,(t) \tag{6}$$

$$d_t w_1(t) = k_{21,r}\, w_1(t) - (k_{12,r} + k_{10} + k_l)\, v_1(t) \tag{7}$$

$$d_t w_2(t) = k_{12,r}\, w_1(t) - k_{21,r}\, w_2(t) \tag{8}$$

with the initial conditions $w_1(0) = 0$ and $w_2(0) = m_0$, where m_0 represents the molecular concentration released from nanonetworks. $k_{12,r}$ and $k_{21,r}$ is reverse equivalents of kinetic constants k_{12}; and k_{21} respectively and k_l is the detachment constant of ligand–receptor pairing.

4. The Proposed Methodology

In the present section, we discuss the methodology employed for anomaly detection using deep learning technologies in bioluminescent interfaces of IoBNT systems. The proposed scheme comprises the following components: (i) *data generation and pre-processing* of bioluminescence parameter values, (ii) *neural network structures*, (iii) *experiment testbed,* and *measurement metrics*. The scheme monitors in vivo bio interface parameters to discriminate and identify anomalous operations. Using in-body parameter profiling allows higher accuracy in the classification of anomalies generated by external attack vectors compared to earlier studies solely relying on in vitro metrics such as bioFET electric voltage and current used [25], which may significantly vary due to electromagnetic interference and/or system faults, resulting in discrepancies. The following sub-sections detail each component of the proposed methodology.

4.1. Data Pre-Processor

The data pre-processor module accounts for collecting and sanitizing data. The parameters employed in the proposed scheme are unique to bio-luminescence-based cyber interface [4]. This work uses a synthetic dataset that consists of 12 key parameters based on existing literature in bioluminescence-based bio–cyber interface proposal for IoBNT [5,20,24]. The usage of synthetic data is a necessity due to the novelty of molecular communication using bio–cyber interfacing in IoBNT devices, which is yet to be fully realized practically. The data for this work depends upon the typical output generated by a bioluminescence-based bio–cyber interface. As described in Section 3, bio–cyber interface consists of two units: an electro-bio transduction unit and a bio-electro transduction unit. Each unit generates an output which is expressed in Equations (2) and (5). The performance of the IoBNT system depends on the output of bio–cyber interface units. Various parameters are required to derive these output equations. The description and value ranges of these parameters are presented in Table 2. These parameters are unique to bioluminescence-based bio–cyber interfaces and inappropriate values of these parameters represent anomalies in the system. The target is to ensure that appropriate values of these parameters are employed in the IoBNT system operation in order to achieve the required output. It is desired to find a set of parameters with minimal elements (features) whose values can be adjusted to ensure efficient communication. These parameters will be used as features to train our neural networking models. In the simulation, values of evaluation parameters used are taken from the literature and scaled up or down in the same order. A brief description of parameters and possible consequences of their abnormal values (out of the given range) are provided in Table 1 and described below:

Table 2. Simulation Parameters.

Ψ_0 is the concentration of message molecules encapsulated in liposomes, injected by an electro-bio transduction unit. A sufficient value of 0 ensures that accurate information is delivered to the targeted site. Tampered or manipulated values of this parameter, in (2), result in malfunctioning of nanomachines at the target site.

k_{10} is the elimination rate that elucidates the loss of information molecules due to the adhesion process, reaction process, absorption by non-targeted sites, phagocytosis, and elimination by the liver. Increased values of this parameter, used in (3), indicate that system might be under blackhole attack and information molecules are attracted by some malicious nanosystems [22,23].

a_M is the Michaelis-Menten constant. This parameter is an essential cofactor for ATP and its intracellular availability is dynamically regulated. It is observed as a weak parameter here as there is little effect of this constant on the bioluminescence intensity compared to other parameters as given in (5).

a_l is the catalytic reaction constant. This parameter aids in adjusting the bioluminescent intensity.

ATP is the concentration of adenosine triphosphate (ATP) and it is an important parameter as its appropriate value contributes to the bioluminescence intensity I0, which turns on the transmitter.

m_0 is the molecular concentration parameter. It should be high enough to turn ON the transmitter and low enough to guarantee that the drug concentration does not reach toxic levels around healthy cells.

λ is the release rate constant and an important parameter. It is necessary to realize complete control of the drug release, from the initial release to a suitable release rate to ensure the best delivery profile for maximizing the treatment benefits while minimizing the amount of drugs in the other parts of the body.

k_l is the ligand–receptor binding constant. k_l is a crucial parameter as it represents whether the ligand–receptor bond happened or not. Inappropriate values of this parameter in (6), might indicate congestion, due to hits of ligand molecules with already busy receptors.

$k_{21,r}$ is the forward rate constant. $k_{21,r}$, and other rate constants i.e., k_{12}, $k_{12,r}$ and k_{21} depends upon the concentration difference in the blood vessel network and nanonetwork, size of endothelial cell separating the blood vessel network and nanonetwork, and properties of the diffusing molecules. A sufficient value of this parameters indicates that more information molecules are received by the bio–cyber interface, resulting in higher bioluminescence.

Data set quality influences deep learning classification results. In the present work, data from simulated in vivo parameters of bioluminescent interfacing summarized in Table 1, will be used using binary classification. Binary classification will require anomalous or normal labeling of generated data during experimentation. Data will be split for testing and training, with a lower proportion of anomalous (attack) data conforming to realistic scenarios and to provide an accurate estimation of classification ability.

4.2. Neural Network Structures

The present work utilizes deep learning methods to analyze anomaly detection capability using bioluminescence intensity parameters of BBI. We first employ the convolutional neural network (CNN) algorithm offering lower latency during training and then offering weightage adjustment after setting initial parameters. The CNN allows the reduction of higher dimensionality using maximum pooling on features and indicates the most suitable feature traits. After testing CNN, recurrent neural network (RNN) structures are used to extract temporal characteristics of data. During extensive validation testing is done to observe the efficiency of different models and to devise a hybrid structure providing maximum classification benefits. The network structures and respective construction are discussed in the following sub-section.

Convolutional model: CNNs offer optimality in feature extraction at relatively lower latency through the creation of associations within data. Traditional implementations use the entry-level input layers to determine as well as learn features while secondary layers can derive more complex characteristics, representing data correlations. Convolution layer kernels filter features, each iteration producing a graph of features provided by each kernel. A pooling layer can be used for feature compression and selection, and the network converges on a final layer allowing connectivity to combine and categorize the model.

The complexity of CNN layers increases with the number of input parameters (exponentially) and can reduce network performance. The error rate is calculated using loss functions. Convolution operation depends on the number of kernels that produce proportional output, the step size determining the amount of data processed in any given internal and the filling factor to smooth out the learning. The convolution operation is given by (9).

$$O_{(I,j)} = \Sigma_{(x=0)}\,\Sigma_{(y=0)}\,K_{(x,y)}\cdot A_{(q-x,j-y)} \tag{9}$$

In total, the present study used six convolution layers, having a kernel size of 4 × 4 providing feature extraction capability. The first CNN layer has a relatively lower number of only twelve neurons for initial data processing. Successive layers provide 64 neurons each to improve classification. A dropout layer is added afterward to deal with overfitting, a common challenge in CNN. The maximum pooling layer provides feature differentiation, while the convolution layer processes the final set of features, compressed, and mapped by the dense layer at the end. A visual description of the employed CNN structure is provided in Figure 5.

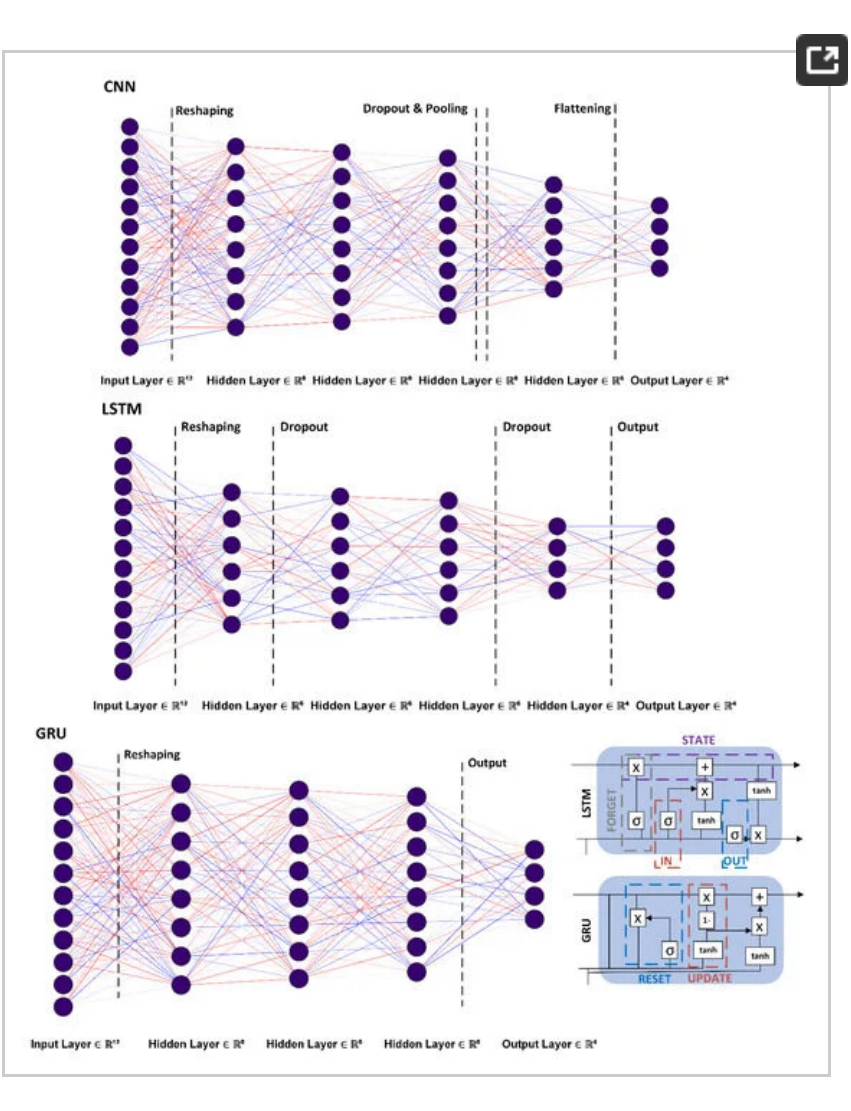

Figure 5. Neural network structures: CNN, LSTM, GRU, and RNN internal cell structure.

Recurrent model: RNN models incorporate interconnected neurons in each layer so that the output of preceding layers is taken into consideration while computing the input to subsequent layers. Stateful structure and feedback mechanism allow RNN to understand temporal characteristics for anomaly detection [39]. RNN however, suffers from the vanishing gradient problem, and therefore, practical applicability on larger data series may require the use of long-short term memory (LSTM) and gated recurrent unit (GRU) implementations. In the present study, we utilize LSTM and GRU models using similar network structures and operational values. Data is input to the first layer, succeeded by the dropout layer adding an element of uncertainty for better training. Models and the individual cell structures are illustrated in Figure 5. Fewer neurons in hidden layers are used in LSTM (36) compared to GRU (64) in the hidden layers, as the former requires a greater number of training parameters and results in higher training latency. The dense layer is the final processing primitive in either model. The *sigmoid* activation is used in GRU and LSTM for filtering and output is sent using a tangential function (*TanH)* [38]. The LSTM and GRU operations are represented by (10)–(13) and (14)–(16) respectively.

Long short-term memory (LSTM) model input and variables.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{10}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_c) \tag{11}$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{12}$$

$$c_t = \tanh(W_C\cdot[h_{t-1}, x_t] + b_C) \tag{13}$$

Gated recurrent unit (GRU) model input and variables.

$$r_t = \mathrm{sigmoid}(W_r[h_{t-1}, x_t] + b_r) \tag{14}$$

$$z_t = \mathrm{sigmoid}(W_r[h_{t-1}, x_t] + b_z) \tag{15}$$

$$r_t = \mathrm{sigmoid}(W_r[r_t.h_{t-1}, x_t] + b_c) \tag{16}$$

The sigmoid, rectified linear unit (ReLU), and general loss functions are given by (17)–(19), respectively.

$$g(z) = 1/[1 + e^{-z}],\ g'(z) = g(z)(1 - g(z)) \tag{17}$$

$$g(z) = \max\,(0, z),\ g(z) = \{1, z > 0, 0 \text{ otherwise} \tag{18}$$

$$L(W) = 1/n\ \textstyle\sum_{k=0\rightarrow n} y^{(i)}\log\,(f(x^{(i)}; W)) + (1 - y^{(i)})\,\log\,(1 - f(x^{(i)}; W)) \tag{19}$$

In (18), (19), $y^{(i)}$ is the actual value while $f(x^{(i)}; W)$ is the prediction.

4.3. Experiment and Measurements

The present section discusses the experiment specifics of the study, detailing data generation, the tools used for implementing deep learning algorithms, and the measurements recorded.

4.3.1. Data Processing

The primary steps involved in *data processing* involve *class derivation and labeling*, *value normalization*, and *encoding* for improved model fitting. An illustration of the main steps is illustrated in Figure 6, and these are discussed in detail as follows.

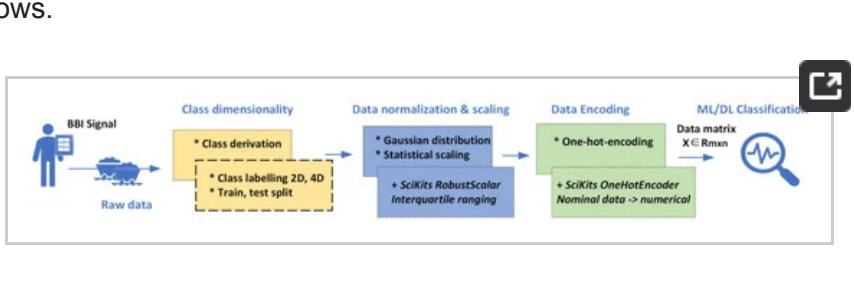

Figure 6. Data processing—primary steps.

Step 1: Class-derivation and labelling: In total 7,985,539 records each having twelve feature values were generated using the bioluminescence parameters given by (1)–(8), and equally split for training and testing summarized earlier in Table 2. As per existing literature, out-of-bound variation in a minimum of three out of twelve (or greater) parameters of any given input vector constitutes an anomaly [2,7,20,33,41]. Therefore, in two-dimensional (2D) class distribution abnormal parameters constituting more than ~25% of an input vector result in final class derivation being *anomalous*, and otherwise *normal*. Using 2D binary distribution, 6,866,701 samples represent *normal* and 1,117,829 *anomalous* vectors, indicating an overall ~86:14 probability split between normal and anomalous data as given in Table 3. However, having uniform 2D class labeling does not completely account for the scarcity of instances where normal in vivo channel communication presents realistic outliers, and the resulting operation is not necessarily anomalous due to realistic variations in molecular traffic [2,7,15,23]. In 4D classification, the normal-anomalous (NA) class realizes the normal operation even though the threshold parameter varies outside the threshold; and the anomalous-normal (AN) class represents abnormal operation despite parameters ranging within the stipulated threshold. NA and AN vector instances constitute ~3.4–3.9% of normal and anomalous vectors based on existing studies [7,8,21,22]. The resulting data is stored in matrix $X \in R_{mxn}$ where *m* represents the number of samples and *n* is the number of variables (twelve). Vector $L \in R_{mx1}$ provides a class label (string).

Table 3. Dataset Specification.

Step 2: Data normalization and scaling: The parameter density distribution of generated data (training and testing) for each of the 4Dclass-labels is provided in Figure 7, and an online dataset representation is provided in [44]. The data follows a Gaussian probability distribution given by (20) and is symmetric about the mean.

$$f(x) = 1/\sigma\ \sqrt{(2\pi)} \times e^{-1/2(x-\mu/\sigma)^{\wedge}2} \tag{20}$$

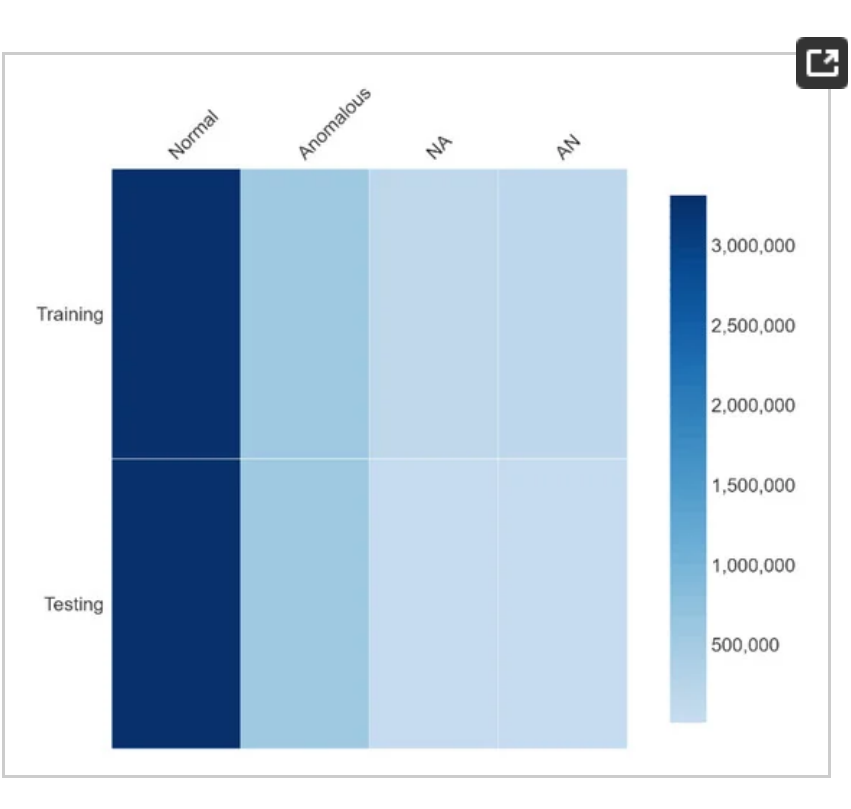

Figure 7. Density distribution—bioluminescence data.

In (20), x represents the class variable being examined, f(x) is the probability function, μ is the mean, and the σ standard deviation. The generated data follows a Gaussian distribution and does not require smoothing filtration [37]. However, normal BBI operations may encompass signal fluctuations, and nonlinear responses resulting in outliers that can affect classification accuracy. Although these outlier values on data distribution edges (NA, AN) have a low occurrence probability, yet capable of over-representation during training. To avoid model skewing inherent relative variability in BBI parameters requires redressal using statistical methods to achieve high throughput during classification. Complementary techniques such as local-neighbor density deviation (LOCAL), isolation using modeling (ISO), and elliptic envelope (ELLI) can cause decision edge confusion, data scattering, and computationally intensive [27,28,37]. Statistical outlier detection using robust scaling is used in the present study to maintain data characteristics while reducing data scattering with minimal overhead. Using the median (50th percentile), and 25th and 75th percentiles, in robust scaling each parameter median is subtracted and divided by the interquartile range (IQR) [32]. This ensures resulting variables conform to zero mean, median, and standard deviation of one while remaining non-skewed by outliers and maintaining relative relationship with other variables (parameters). Outliers depicting normal class with MC channel variation are, therefore, upheld and contribute to the discriminatory ability of the resulting classifier in recognizing false positives, as well as negatives. Robust scaling for the present experiment is achieved using Python SciKits RobustScalar class, setting with_centring argument centered to zero, and with_scaling set to IQR, and both values defaulting to True, as listed in Table 4 [37]. Given the stochastic ability of algorithms differences in numerical precision are possible in iterative runs during classification and as such the percentile ranges for scaling were varied (20–80, and 30–70) to derive optimal accuracy. Compared to similar studies, the proposed data normalization technique fits well on low-dimensional tall data offering minimum scattering while exploring full (traffic) characteristics for optimal classification [28,29].

Table 4. System Variables.

Step 3: Data encoding: The data composition matrix ($X \in R_{mxn}$) comprises twelve continuous data parameters (n = 12) with range-bound probable values, as well as one nominal variable (L) representing 4D class labels without any intrinsic ordering. One hot encoding using SciKits OneHotEncoder class is a used for numerical conversion of this nominal data before the application of classification algorithms [37]. Since the class labeling does not have any intrinsic order, hence ascending encoding (0–3) cannot be used otherwise ML algorithms will give priority to higher numbers. Each category is hence, transformed into a new column and given a binary range (0,1) depending on the occurrence of the respective label. Post-transformation dataset, therefore, results in a greater number of features due to one hot encoding of traffic class labeling. Compared to other nominal conversion schemes such as Bag-of-Word (BoW), and MinMaxScaling, one hot encoding provides a single-stage conversion suitable for BBI parameters relying on substantially lower computing capability favored in IoT environments [29,31]. The pre-processed dataset is subsequently input to neural algorithms using control variables described in the next sub-section.

4.3.2. Neural Network Structures and Control Variables

The neural network structures tested include CNN, LSTM, and GRU. The model structures used are provided in Figure 5. The CNN model comprised a single input layer containing 12 neurons, four hidden convolutional layers each containing 64 neurons, and a maximum pooling, dropout, and flattening layer. LSTM also comprises one input layer (16 neurons) as well as a dropout in hidden layers. The GRU model has an input layer (12 neurons), and three GRU layers having a greater number of neurons (64, 128, 156) and dropout in hidden layers. As discussed earlier, GRU has an overall greater number of neurons in comparison with LSTM models, but the operational values used are smaller. The dropout for all models was set to the commonly used value of 0.5. For activation, ReLU was employed as preferred in earlier studies [34,37]. The output layer uses federate AI (FATE) [45] and is classified using softmax activation [37]. Control variables used for experimentation are listed in Table 4. Training and testing on each neural network model are conducted using an epoch of 10 on a trial-and-error basis with the same input–output dimensions and class (category) labeling. To reduce overall cost-function Nadam-Nesterov-acceleration adaptive moment estimation (gradient descent) is employed using Keras [46]. Nadam optimizes model learning and classification by summative exponential reduction of floating averages on preceding and successive gradient values. For computing the overall loss function, seeking to minimize the training phase, categorical cross entropy available in Keras is used [47]. The experiment was carried out on a researcher's machine with an Intel Core i7-960 16 MB chipset at 3.2 GHz with RAM 64 GB. The operating system used was Ubuntu 18.04.5 (64-bit) hosting a testing platform using Keras 2.4.0 and TensorFlow library 2.8.0. [48]. The designed neural network models were evaluated on the generated dataset to record accuracy, precision, recall, as well as the false positive and negative rates given in Table 5.

Table 5. Metric Rubric.

5. Results and Discussion

The present section discusses the evaluation results of proposed deep learning structures from a quantitative and qualitative perspective as well as an analysis of the optimum classification model.

5.1. Quantitative Results

The recorded efficiency of each model for four-dimensional parameter classification using the network structures and system variables discussed earlier is given in Table 6. The quantitative results are systematically analyzed to determine the best classifier from an operational perspective considering training latency, testing duration, as well as reported accuracy and precision. CNN is considered first, with additional structures subsequently added to the testbed for evaluation. For comparison purposes, a 1D shallow CNN is also considered.

Table 6. Result Evaluation—Quantitative.

The time complexity of each neural network structure is considered to indicate the time involved in training and testing the classifier. Each classification model was considered under ten successive training iterations. The shallow network outperforms all others in terms of latency which is expected. Each training iteration of the shallow neural network (1D-CNN) was recorded at approximately 2.7 s having a total training time of 26.81 s. Deep-layered CNN training time was the lowest among all remaining non-hybrid and hybrid classifiers at 215.89 s. CNN processed 399,226 samples for each training iteration in ~21.58 s. The LSTM consumed 78.56 s and GRU 92.12 s for each training iteration recording a total training time of 785 s and 921 s respectively. Among the hybrid models, the CNN + LSTM model offered the advantage of faster extraction of relevant features (due to CNN) and a higher cross-dimensionality analysis (provided by LSTM) at a relatively lower training time of 315.6 s compared to the CNN + GRU model at 468.12 s. During testing and validation, the 3,992,271 samples were completed in ~117 s by CNN, ~257 s by LSTM, and ~312.89 s by GRU. Similarly, the CNN + LSTM consumed around 129.31 s during testing compared to the CNN + GRU, which recorded 221 s.

ccuracy, precision, recall, and f1-score for each of the neural network models is demonstrated in Figure 8. ng non-hybrid classifiers, the CNN reported highest accuracy (91.56%) and precision (92.21%) when compared